

Bayesian Robustness to Outliers in Linear Regression and Ratio Estimation

Alain Desgagné^a, Philippe Gagnon^{b,*}

^a*Département de mathématiques, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3P8*

^b*Département de mathématiques et statistique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J7*

Abstract

Whole robustness is a nice property to have for statistical models. It implies that the impact of outliers gradually decreases to nothing as they converge towards plus or minus infinity. So far, the Bayesian literature provides results that ensure whole robustness for the location-scale model. In this paper, we make two contributions. First, we generalise the results to attain whole robustness in simple linear regression through the origin, which is a necessary step towards results for general linear regression models. We however allow the variance of the error term to depend on the explanatory variable. This flexibility leads to the second contribution: we provide a simple Bayesian approach to robustly estimate finite population means and ratios. The strategy to attain whole robustness is simple since it lies in replacing the traditional normal assumption on the error term by a super heavy-tailed distribution assumption. As a result, users can estimate the parameters as usual, using the posterior distribution.

Keywords: Built-in Robustness, Simple Linear Regression, Ratio Estimator, Finite Populations, Population Means, Super Heavy-Tailed Distributions

1. Introduction

Conflicting sources of information may contaminate the inference arising from statistical analysis. The conflicting information may come from outliers and also prior misspecification. In this paper, we focus on robustness with respect to outliers in a Bayesian simple linear regression model through the origin.

We say that a conflict occurs when a group of observations produces a rather different inference than that proposed by the bulk of the data and the prior. Light-tailed distribution assumptions on the error term can lead to an undesirable compromise where the posterior distribution concentrates on an area that is not supported by any source of information. We believe that the appropriate way

*Corresponding author

Email addresses: `desgagne.alain@uqam.ca` (Alain Desgagné), `gagnonp@dms.umontreal.ca` (Philippe Gagnon)

to address the problem is to limit the influence of outliers in order to obtain conclusions consistent with the majority of the observations.

[Box and Tiao \(1968\)](#) were the first to introduce a robust Bayesian linear regression model. They proposed to assume that the distribution of the error term is a mixture of two normals with one component for the nonoutliers and the other one, with a larger variance, for the outliers. This approach has been generalised by [West \(1984\)](#) who modelled errors with heavy-tailed distributions constructed as scale mixtures of normals, which includes the Student distribution. More recently, [Peña et al. \(2009\)](#) introduced a different robust Bayesian method where each observation has a weight decreasing with the distance between this observation and most of the data. They proved that the Kullback-Leibler divergence from the posterior arising from the nonoutliers only to the posterior arising from the sample containing outliers is bounded.

So far, the literature only provides solutions to attain whole robustness for the estimation of the slope in the model of regression through the origin (e.g. if we assume that the error term has a Student distribution instead of a normal, see the results of [Andrade and O'Hagan \(2011\)](#) in a context of location-scale model); however, only partial robustness is reached for the estimation of the scale parameter of the error term. Note that partial robustness means that the outliers have a significant but limited influence on the inference, as the conflict grows infinitely. In this paper, we go a step further: we attain whole robustness to outliers for both the slope and the scale parameters, in the sense that the impact of outliers gradually decreases to nothing as they converge towards plus or minus infinity. In order to achieve this, we generalise the results of [Desgagné \(2015\)](#), which ensure whole robustness for both parameters of the location-scale model simultaneously, to the simple linear regression model through the origin. Our work is thus aligned with the *theory of conflict resolution in Bayesian statistics*, as described by [O'Hagan and Pericchi \(2012\)](#) in their extensive literature review on that topic.

The strategy to attain robustness is, instead of assuming the traditional normality of the errors in the model, to assume that they have a super heavy-tailed distribution. The general model (with no specific distribution assumption on the error term) is described in Sect. 2.1. The class of super heavy-tailed distributions that we consider, which are log-regularly varying distributions, is presented in Sect. 2.2. When assuming a super heavy-tailed distribution on the error term, the resulting model is characterised by its built-in robustness that resolves conflicts in a sensitive and automatic way, as stated in our robustness results given in Sect. 2.3. The main result is: the convergence of the posterior distribution towards the posterior arising from the nonoutliers only, when the outliers approach plus or minus infinity. Although our results are Bayesian analysis-oriented, they reach beyond this paradigm through the robustness of the likelihood function, and therefore, of both slope and scale maximum likelihood parameter estimation. Note that the generalisation of the results described in this paper to the simple linear regression (with an unknown intercept) is not trivial.

Linear regression through the origin is particularly useful for the estimation of finite population means and ratios. In Sect. 3, we establish the link between this linear regression model and finite population sampling. We also present real-life situations of finite population sampling in which our approach is useful, and a simulation study. Our approach is compared with the nonrobust (with the normal assumption) and partially robust (with the Student distribution assumption) approaches. It is showed that our model performs as well as the nonrobust and the partially robust models in

absence of outliers, in addition to being completely robust. It indicates that, by only changing the assumption on the error term, we obtain adequate estimates in absence or presence of outliers. These estimates are computed as usual from the posterior distribution.

2. Resolution of Conflicts in Simple Linear Regression Through the Origin

2.1. Model

- (i) Let $Y_1, \dots, Y_n \in \mathbb{R}$ be n random variables and $x_1, \dots, x_n \in \mathbb{R} \setminus \{0\}$ be n known constants, where $n > 2$ is assumed to be known. We assume that

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, \dots, n,$$

where $\epsilon_1, \dots, \epsilon_n, \beta \in \mathbb{R}$ are $n + 1$ conditionally independent random variables given $\sigma > 0$ (a random variable) with a conditional density for ϵ_i given by

$$\epsilon_i | \beta, \sigma \stackrel{\mathcal{D}}{=} \epsilon_i | \sigma \stackrel{\mathcal{D}}{\sim} \frac{1}{\sigma |x_i|^\theta} f\left(\frac{\epsilon_i}{\sigma |x_i|^\theta}\right), \quad i = 1, \dots, n,$$

$\theta \in \mathbb{R}$ being a known constant.

- (ii) We assume that f is a strictly positive continuous probability density function on \mathbb{R} that is symmetric with respect to the origin, and that is such that both tails of $|z|f(z)$ are monotonic, which implies that the tails of $f(z)$ are also monotonic. The density f can have parameters, e.g. a shape parameter; however, their value is assumed to be known.
- (iii) We assume that the prior of β and σ , given by $\pi(\beta, \sigma)$, is non-negative and that $\min(\sigma, 1)\pi(\beta, \sigma)$ is bounded (or equivalently that $\pi(\beta, \sigma)/\max(1, 1/\sigma)$ is bounded). Note that, if we have no prior information, we can set $\pi(\beta, \sigma) \propto 1/\sigma$, the usual non-informative prior for this type of random variables, or simply $\pi(\beta, \sigma) \propto 1$.

From this perspective, x_1, \dots, x_n represent observations of the explanatory variable, the dependent variable and the error term are respectively represented by the continuous random variables Y_1, \dots, Y_n and $\epsilon_1, \dots, \epsilon_n$, and the parameter β represents the slope of the regression line. Note that no assumptions are made on the explanatory variable, except that the value 0 cannot be observed; it can be continuous, discrete, with any distribution.

The scale of the distribution of the error term is $\sigma |x_i|^\theta$ and, therefore, the variability of the errors increases (decreases) as x_i moves away from 0 when $\theta > 0$ ($\theta < 0$). This model can thus be used in a context of heteroscedasticity and, when the classical framework is considered, i.e. a frequentist setting with the assumption that f is the standard normal density, $\sigma |x_i|^\theta$ also represents the standard deviation of the error ϵ_i . In this situation, the maximum likelihood estimator of β is the weighted average of the y_i/x_i given by $\hat{\beta} = \sum_{i=1}^n w_i(y_i/x_i)$, where $w_i = |x_i|^{2(1-\theta)} / \sum_{j=1}^n |x_j|^{2(1-\theta)}$.

An important drawback of the classical framework is that outliers have a significant impact on the estimation, due to the normal assumption. In this paper, we study robustness of the estimation

of β and σ . The objective is to find sufficient conditions to attain whole robustness, meaning a gradual decrease in the impact of outliers as they converge towards plus or minus infinity, to ultimately reach a level where their impact is null. The nature of the results presented in Sect. 2.3 is asymptotic, in the sense that some y_i 's approach $+\infty$ or $-\infty$. The known vector $\mathbf{x}_n := (x_1, \dots, x_n)$ is considered as fixed. In Sect. 3.1, we explain that studying this theoretical framework is sufficient to attain, in practise, robustness against any type of outliers (i.e. outliers because of their extreme x value, extreme y value, or both).

Among the n observations of Y_1, \dots, Y_n , denoted by \mathbf{y}_n , we assume that $k > 2$ of them, denoted by the vector \mathbf{y}_k , form a group of nonoutlying observations, m of them are considered as “negative slope outliers”, with relatively small (large) values of y_i when x_i is positive (negative), and p of them are considered as “positive slope outliers”, with relatively large (small) values of y_i when x_i is positive (negative), with $k + m + p = n$. Note that we use the letter m for “minus” because the related outliers attract the slope towards negative values, and analogously, we use the letter p for “positive”. For $i = 1, \dots, n$, we define the binary functions k_i, m_i and p_i as follows: if y_i is a nonoutlying value, $k_i = 1$; if it is a negative slope outlier, $m_i = 1$ and if it is a positive slope outlier, $p_i = 1$. These functions take the value of 0 otherwise. Therefore, we have $k_i + m_i + p_i = 1$ for $i = 1, \dots, n$, with $\sum_{i=1}^n k_i = k$, $\sum_{i=1}^n m_i = m$ and $\sum_{i=1}^n p_i = p$. We assume that each outlier converges towards $-\infty$ or $+\infty$ at its own specific rate, to the extent that the ratio of two outliers is bounded. More precisely, we assume that $y_i = a_i + b_i\omega$, for $i = 1, \dots, n$, where a_i and b_i are constants such that $a_i \in \mathbb{R}$ and

- (i) $b_i = 0$ if $k_i = 1$,
- (ii) $b_i < 0$ if y_i is “small”, that is if $x_i < 0, p_i = 1$ or $x_i > 0, m_i = 1$,
- (iii) $b_i > 0$ if y_i is “large”, that is if $x_i < 0, m_i = 1$ or $x_i > 0, p_i = 1$,

and we let $\omega \rightarrow \infty$.

Let the joint posterior density of β and σ be denoted by $\pi(\beta, \sigma \mid \mathbf{y}_n)$ and the marginal density of (Y_1, \dots, Y_n) be denoted by $m(\mathbf{y}_n)$, where

$$\pi(\beta, \sigma \mid \mathbf{y}_n) = [m(\mathbf{y}_n)]^{-1} \pi(\beta, \sigma) \prod_{i=1}^n \frac{1}{\sigma |x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right), \quad \beta \in \mathbb{R}, \sigma > 0.$$

Let the joint posterior density of β and σ arising from the nonoutlying observations only be denoted by $\pi(\beta, \sigma \mid \mathbf{y}_k)$ and the corresponding marginal density be denoted by $m(\mathbf{y}_k)$, where

$$\pi(\beta, \sigma \mid \mathbf{y}_k) = [m(\mathbf{y}_k)]^{-1} \pi(\beta, \sigma) \prod_{i=1}^n \left[\frac{1}{\sigma |x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right) \right]^{k_i}, \quad \beta \in \mathbb{R}, \sigma > 0.$$

Note that if the prior $\pi(\beta, \sigma)$ is proportional to 1, the likelihood functions, given by the product term in the posteriors above, can also be expressed as follows:

$$\mathcal{L}(\beta, \sigma \mid \mathbf{y}_n) = m(\mathbf{y}_n) \pi(\beta, \sigma \mid \mathbf{y}_n) \quad \text{and} \quad \mathcal{L}(\beta, \sigma \mid \mathbf{y}_k) = m(\mathbf{y}_k) \pi(\beta, \sigma \mid \mathbf{y}_k). \quad (1)$$

Proposition 1. *Considering the Bayesian context given in Sect. 2.1, the joint posterior densities $\pi(\beta, \sigma \mid \mathbf{y}_k)$ and $\pi(\beta, \sigma \mid \mathbf{y}_n)$ are proper.*

The proof of Proposition 1 can be found in the supplementary material.

2.2. Log-Regularly Varying Distributions

As mentioned in the introduction, our approach to attain robustness is to replace the traditional normal assumption on the error term by a log-regularly varying distribution assumption. The definition of such a distribution is now presented.

Definition 1 (Log-regularly varying distribution). *A random variable Z with a symmetric density $f(z)$ is said to have a log-regularly varying distribution with index $\rho \geq 1$ if $zf(z) \in L_\rho(\infty)$, meaning that $zf(z)$ is log-regularly varying at ∞ with index $\rho \geq 1$.*

Log-regularly varying functions is an interesting class of functions with useful properties for robustness. By definition, they are such that $g \in L_\rho(\infty)$ if $g(z^\nu)/g(z)$ converges towards $\nu^{-\rho}$ uniformly in any set $\nu \in [1/\tau, \tau]$ (for any $\tau \geq 1$) as $z \rightarrow \infty$, where $\rho \in \mathbb{R}$. This implies that for any $\rho \in \mathbb{R}$, we have $g \in L_\rho(\infty)$ if and only if there exists a constant $A > 1$ and a function $s \in L_0(\infty)$ (which is called a log-slowly varying function) such that for $z \geq A$, g can be written as $g(z) = (\log z)^{-\rho} s(z)$. It gives you an overview of the tail behaviour of log-regularly varying distributions. Note that an example of such a distribution is presented in Sect. 3.1. For more information on log-regularly varying distributions, we refer the reader to [Desgagné \(2013\)](#) and [Desgagné \(2015\)](#).

2.3. Resolution of conflicts

The results of robustness are now given in Theorem 1.

Theorem 1. *Consider the model and the context described in Sect. 2.1. If we assume that*

- (i) $zf(z) \in L_\rho(\infty)$, with $\rho \geq 1$ (i.e. that f is a log-regularly varying distribution),
- (ii) $k > \max(m, p)$ (i.e. that both the negative and positive slope outliers are fewer than the nonoutliers),

then we obtain the following results:

(a)

$$\lim_{\omega \rightarrow \infty} \frac{m(\mathbf{y}_n)}{\prod_{i=1}^n [f(y_i)]^{m_i + p_i}} = m(\mathbf{y}_k),$$

(b)

$$\lim_{\omega \rightarrow \infty} \pi(\beta, \sigma \mid \mathbf{y}_n) = \pi(\beta, \sigma \mid \mathbf{y}_k),$$

uniformly on $(\beta, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$, for any $\lambda \geq 0$ and $\tau \geq 1$,

(c)

$$\lim_{\omega \rightarrow \infty} \int_0^\infty \int_{-\infty}^\infty |\pi(\beta, \sigma | \mathbf{y}_n) - \pi(\beta, \sigma | \mathbf{y}_k)| d\beta d\sigma = 0,$$

(d) As $\omega \rightarrow \infty$,

$$\beta, \sigma | \mathbf{y}_n \xrightarrow{\mathcal{D}} \beta, \sigma | \mathbf{y}_k,$$

and in particular

$$\beta | \mathbf{y}_n \xrightarrow{\mathcal{D}} \beta | \mathbf{y}_k \quad \text{and} \quad \sigma | \mathbf{y}_n \xrightarrow{\mathcal{D}} \sigma | \mathbf{y}_k,$$

(e)

$$\lim_{\omega \rightarrow \infty} [m(\mathbf{y}_k)/m(\mathbf{y}_n)] \mathcal{L}(\beta, \sigma | \mathbf{y}_n) = \mathcal{L}(\beta, \sigma | \mathbf{y}_k),$$

uniformly on $(\beta, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$, for any $\lambda \geq 0$ and $\tau \geq 1$.

The proof of Theorem 1 can be found in the supplementary material. Note that, when $x_1 = \dots = x_n = 1$, the simple linear regression model through the origin becomes the location-scale model, and this highlights the fact that our results generalise those of [Desgagné \(2015\)](#).

Theorem 1 is particularly appealing for its simplicity, and therefore, for its practical use. Indeed, condition (i) only indicates that modelling must be done using a density f with sufficiently heavy tails, specifically with a log-regularly varying distribution (see Definition 1). For that purpose, [Desgagné \(2015\)](#) introduced the family of log-Pareto-tailed symmetric distributions, which belongs to the family of log-regularly varying distributions and therefore satisfies condition (i). A special case of log-Pareto-tailed symmetric distributions, called the log-Pareto-tailed standard normal distribution with parameters $\alpha > 1$ and $\phi > 1$, is given in Sect. 3.1. It exactly matches the standard normal on the interval $[-\alpha, \alpha]$, with log-Pareto tails that behave like $(1/|z|)(\log |z|)^{-\phi}$. Note that condition (i) involves only the tails of the function $|z|f(z)$.

Condition (ii) indicates that both the negative and positive slope outliers must be fewer than the nonoutlying observations, i.e. $m < k$ and $p < k$. In other words, the nonoutlying observations must form the largest group. For instance, with a sample of size $n = 25$, the model rejects up to 16 outliers if they are split in $m = 8$ negative and $p = 8$ positive slope outliers, which leaves $k = 9$ nonoutliers. At the other end of the spectrum, in the situation where all the outliers are of the same type, for instance all positive slope outliers (which implies that $m = 0$), the model rejects up to $p = 12$ outliers, which leaves $k = 13$ nonoutliers. Numerical simulations seem to confirm our expectation that a larger difference between k and $\max(m, p)$ results in a more rapid rejection of the outliers.

Not only do the conditions of Theorem 1 are simple and intuitive, the results are also easy to interpret. The asymptotic behaviour of the marginal $m(\mathbf{y}_n)$ is described by result (a). While this result is more of theoretical interest, it is the cornerstone of this robustness theory; it leads to results (b) to (e), which are more practical. Result (b) indicates that the posterior density, arising from the whole sample, converges towards the posterior density arising from the nonoutliers only,

uniformly in any set $(\beta, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$. The impact of the outliers then decreases gradually to nothing as they approach plus or minus infinity.

Result (b) leads to result (c): the convergence in L_1 of the posterior density, arising from the whole sample, towards the posterior density arising from the nonoutlying observations only. This last result implies the following convergence: $\mathbb{P}(\beta, \sigma \in E \mid \mathbf{y}_n) \rightarrow \mathbb{P}(\beta, \sigma \in E \mid \mathbf{y}_k)$ as $\omega \rightarrow \infty$, uniformly for all rectangles $E \in \mathbb{R} \times \mathbb{R}^+$. This result is slightly stronger than convergence in distribution (result (d)) which requires only pointwise convergence. Then, the convergence of the posterior marginal distributions is directly obtained. Therefore, any estimation of β and σ based on posterior quantiles (e.g. posterior medians and Bayesian credible intervals) is robust to outliers. Note that results (a) to (d) are also valid if we assume that $n \geq 2, k \geq 2$ (instead of $n > 2, k > 2$), provided that we assume that $\sigma\pi(\beta, \sigma)$ is bounded (instead of $\min(\sigma, 1)\pi(\beta, \sigma)$ is bounded).

Result (e) indicates that, for a given sample, the likelihood (up to a multiplicative constant that does not depend on β and σ) converges to the likelihood arising from the nonoutliers only, uniformly in any set $(\beta, \sigma) \in E$, where $E = [-\lambda, \lambda] \times [1/\tau, \tau]$. Consequently, the maximum of $\mathcal{L}(\beta, \sigma \mid \mathbf{y}_n)$ thus converges to the maximum of $\mathcal{L}(\beta, \sigma \mid \mathbf{y}_k)$ on the set E and, therefore the maximum likelihood estimate also converges, as $\omega \rightarrow \infty$. Note that, using results (b) to (d), we know that, for both $\pi(\beta, \sigma \mid \mathbf{y}_k)$ and $\pi(\beta, \sigma \mid \mathbf{y}_n)$, the volume on E^c over the volume on E converges to 0 as λ and τ increase; this relation holds in particular if $\pi(\beta, \sigma) \propto 1$ and, in this case, the posterior is proportional to the likelihood.

3. Finite Population Means and Ratios

In order to use the model described in Sect. 2.1, users have to set the value of θ . Different particular values lead to interesting special cases. For instance, when $\theta = 0$, the resulting model is the classical homoscedastic model, with $\text{Var}(\epsilon_i) = \sigma^2$ and $\hat{\beta} = \sum_{i=1}^n x_i y_i / \sum_{j=1}^n x_j^2$, considering the classical framework. When $\theta = 1$, the estimator of β is the unweighted mean of the y_i/x_i , that is $\hat{\beta} = (1/n) \sum_{i=1}^n y_i/x_i$. Probably the most interesting special case results of $\theta = 1/2$ and $x_i > 0$ for all i . Indeed, considering again the classical framework, the estimator of β is $\hat{\beta} = \sum_{i=1}^n y_i / \sum_{i=1}^n x_i$, which is commonly used to estimate the following finite population ratio: $\sum_{i=1}^N y_i / \sum_{i=1}^N x_i$, where y_i and x_i are respectively measures of the variable of interest and of the auxiliary variable on unit i , and N is the population size. The estimator $\hat{\beta} = \sum_{i=1}^n y_i / \sum_{i=1}^n x_i$ is also used to estimate the finite population mean μ_y of a variable of interest y using auxiliary information of a variable x as follows: $\hat{\mu}_y = \hat{\beta} \times \mu_x$, where μ_x is the known population mean of x . This last estimator is known as the ratio estimator and to be more accurate than the simple location model when the variable of interest is correlated with the auxiliary variable. Therefore, robust estimators of β lead to robust estimators of finite population means and ratios. To our knowledge, [Gwet and Rivest \(1992\)](#) introduced the first frequentist outlier resistant alternatives to the ratio estimator, using well known M - ([Huber \(1973\)](#)) and GM - ([Mallows \(1975\)](#)) estimators. Their research was inspired by the work of [Chambers \(1986\)](#), the first author to use regression M -estimators in survey sampling.

In Sect. 3.1, we present real-life situations in which ratio estimation is useful, while illustrating the theoretical results of Theorem 1. First, in a context of estimation of personal disposable income (PDI) per capita, we show that, when we artificially move an observation, its impact on the estimation grows until it reaches a certain threshold. Beyond this threshold, the impact decreases

to nothing as the observation converges towards plus or minus infinity. Second, a more traditional Bayesian analysis is made, in which we study the proportion of income spent on food. More precisely, we present the posterior distributions, with particular emphasis on the impact of outliers, and we compute various estimates from the posteriors. In Sect. 3.2, again in a context of finite population sampling, a simulation study is conducted to compare the accuracy of the estimates arising from our model with those of the nonrobust (the model with the normal assumption) and partially robust (the model with the Student distribution assumption) models.

3.1. Illustration of the Results of Theorem 1

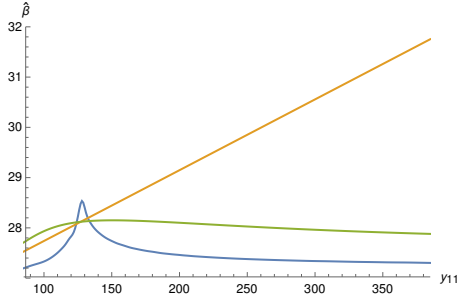
In the first context, we are interested in the estimation of PDI per capita when the available data are the total disposable income (y_i) for n households (in this analysis $n = 20$), and the number of individuals (x_i) in each of these households. Data are presented in Table 1. The PDI per capita, which is a population mean per individual, would be directly computed by $\sum_{i=1}^N y_i / \sum_{i=1}^N x_i$ (where N is the number of households in the population) if the information was available for all the households. We therefore use the simple linear regression model through the origin with $\theta = 1/2$ to estimate this ratio (see Sect. 2.1 for details about the model).

y_i	20.8	9.6	38.6	74.1	108.8	98.7	44.8	77.2	93.2	107.2
x_i	1.0	1.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
y_i	y_{11}	93.6	113.7	123.5	93.5	148.1	147.1	154.0	149.5	173.5
x_i	3.0	4.0	4.0	4.0	4.0	5.0	5.0	5.0	6.0	6.0

Table 1: Total disposable income for household i in thousands of dollars (y_i) and the number of individuals in household i (x_i), for $i = 1, \dots, 20$

In order to illustrate the threshold feature, an observation is randomly chosen (in this analysis, it is the 11th observation), and y_{11} is gradually moved from the value 85 (a nonoutlier) to 385 (a large outlier), while $x_{11} = 3$ remains fixed. The parameters β and σ are estimated for each data set related to a different value of y_{11} using maximum *a posteriori* probability (MAP) estimation with a prior proportional to 1 (which corresponds to maximum likelihood estimation). This process is performed under three models, each corresponding to a different assumption on f : a standard normal density (in this case, $\hat{\beta} = \sum_{i=1}^{20} y_i / \sum_{i=1}^{20} x_i$, the classical ratio estimator), a Student density (the partially robust model) or a log-Pareto-tailed standard normal density (our robust model). The results are presented in Figure 1.

Estimation of the PDI per capita (β) under various distribution assumptions when y_{11} increases from 85 to 385



Estimation of σ under various distribution assumptions when y_{11} increases from 85 to 385

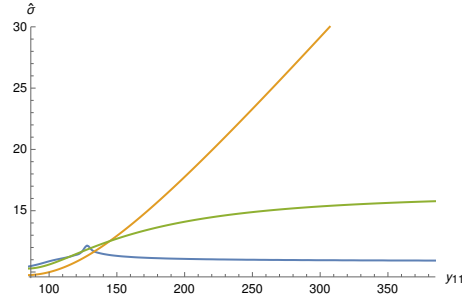


Figure 1: Estimation of the PDI per capita (β) and σ when y_{11} increases from 85 to 385 under three different assumptions on f : standard normal density (orange line), Student density (green line) and log-Pareto-tailed standard normal density (blue line)

The inference is clearly not robust when it is assumed that the error has a normal distribution (orange line) since the values of the point estimates of β and σ increase with y_{11} . Regarding the second model, the degrees of freedom of the heavy-tailed Student distribution have been arbitrarily set to 10 and a known scale parameter of 0.88 has been added to this distribution in order to have the same 2.5th and 97.5th percentiles as the standard normal. The estimation of β is robust as the impact of the outlier slowly decreases after a certain threshold. However, the estimation of σ is only partially robust, i.e. the impact of the outlier is limited, but does not decrease when the outlying value increases. For the last model, we have arbitrarily set the parameter α of the log-Pareto-tailed standard normal distribution to $\alpha = 1.96$, and, according to the procedure described in Section 4 of [Desgagné \(2015\)](#), $\phi = 4.08$ (this procedure ensures that f is continuous and a probability density function). Therefore, all three distributions studied in this section have 95% of their mass in the interval $[-1.96, 1.96]$. The density of the log-Pareto-tailed standard normal distribution is given by

$$f(x) = \begin{cases} (2\pi)^{-1/2} \exp(-x^2/2) & \text{if } |x| \leq \alpha \text{ (the standard normal part),} \\ (2\pi)^{-1/2} \exp(-\alpha^2/2)(\alpha/|x|)(\log \alpha / \log |x|)^\phi & \text{if } |x| > \alpha \text{ (the log-Pareto tails),} \end{cases} \quad (2)$$

and depicted in Figure 2.

Standard Normal Vs Log-Pareto-Tailed Standard Normal

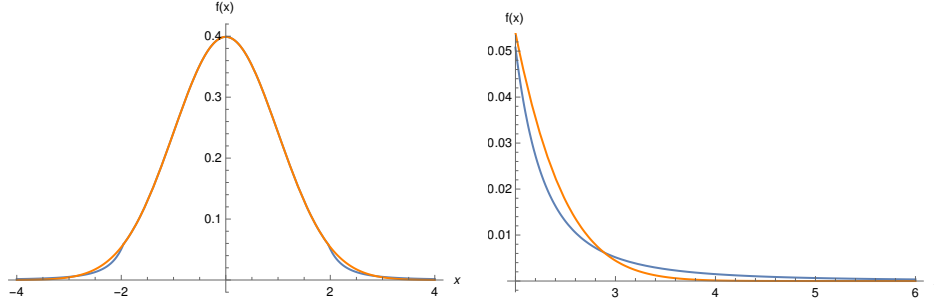


Figure 2: Densities of the standard normal (orange line) and of the log-Pareto-tailed standard normal with $\alpha = 1.96$ and $\phi = 4.08$ (blue line)

For our robust model, it can be seen that y_{11} has an increasing impact on the estimation until this observation reaches a threshold. In this analysis, the threshold is around $y_{11} = 128$, and based on the data set with $y_{11} = 128$, $\hat{\beta} = 28.6$ and $\hat{\sigma} = 12.4$, which is interpreted as: the personal disposable income per capita is approximately 28,600. Beyond this threshold, the impact of the outlier gradually decreases to nothing as the conflict grows infinitely. The point estimates converge towards 27.1 for β and 10.6 for σ , which are the point estimates when (x_{11}, y_{11}) is excluded from the sample. Whole robustness is therefore attained for both β and σ . Note that an increase in the value of the parameter α would result in an increase in the value of the threshold. Setting $\alpha = 1.96$ seems to be suitable for practical use.

In the second context, we are interested in the estimation of the proportion of weekly income spent on food for a population, when data are available per household. If the information was available for the population, we would directly compute the proportion by $\sum_{i=1}^N y_i / \sum_{i=1}^N x_i$, where N is the number of households in the population, and y_i and x_i are respectively the weekly expenditure on food and the weekly income, for household i . This ratio can thus be approximated using the simple linear regression through the origin with $\theta = 1/2$, and again, we compare our robust model with the nonrobust and partially robust models. We use the same Student and log-Pareto-tailed standard normal distributions as in the first context above, but we set the prior $\pi(\beta, \sigma) \propto 1/\sigma$.

Note that the ratio $\sum_{i=1}^N y_i / \sum_{i=1}^N x_i$ can be viewed as the following weighted average: $\sum_{i=1}^N w_i (y_i / x_i)$, where $w_i := x_i / \sum_{i=1}^N x_i$. It means that proportion of weekly income spent on food for a population $\sum_{i=1}^N y_i / \sum_{i=1}^N x_i$ is also a weighted average of proportions of weekly income spent on food per household, where the weight is proportional to the weekly income.

The data set, comprised of the weekly expenditures on food for twenty households and the weekly incomes for these households, is presented in Table 2 and depicted in Figure 3 (a). The posterior distributions of β and σ are presented in Figures 3 (b) and (c). The posterior medians of β are 0.283, 0.306 and 0.319 with 95% highest posterior density (HPD) intervals of (0.217, 0.349), (0.243, 0.366) and (0.240, 0.376) for the nonrobust, partially robust and robust models, respectively. As a result, the proportion of weekly income spent on food for this population is estimated at 0.319 (considering our robust model) with a 95% HPD interval of (0.240, 0.376). The average weekly household expenditure on food of this population can also be estimated using the ratio estimator. Considering our robust model, it is estimated at $\hat{\mu}_y = \hat{\beta} \times \mu_x = 0.319 \times 210 = 66.99$

(considering an average weekly household income of 210 for this population) with a 95% HPD interval of (50.40, 78.96). The posterior medians of σ are 2.183, 2.033 and 1.634 with 95% HPD intervals of (1.559, 3.006), (1.319, 2.960) and (0.961, 2.671), for the nonrobust, partially robust and robust models, respectively.

We observe the presence of two clear outliers: $(x_{17}, y_{17}) = (250.2, 6.1)$ (because of its extremely low y value) and $(x_{20}, y_{20}) = (696.4, 41.1)$ (because of its extremely high x value). In order to draw conclusions based on the bulk of the data and to evaluate the impact of outliers, we redo the analysis while excluding these two outliers. The results are presented in Figure 4. The posterior medians of β are 0.342, 0.339 and 0.343 with 95% HPD intervals of (0.302, 0.382), (0.298, 0.381) and (0.304, 0.382) for the nonrobust, partially robust and robust models, respectively. Therefore, the proportion of weekly income spent on food for this population is estimated at 0.343 (considering our robust model) with a 95% HPD interval of (0.304, 0.382), based on the bulk of the data. Considering our robust model, the average weekly household expenditure on food is now estimated at $\hat{\mu}_y = \hat{\beta} \times \mu_x = 0.343 \times 210 = 72.03$ (considering an average weekly household income of 210 for this population) with a 95% HPD interval of (63.84, 80.22), using the ratio estimator. The posterior medians of σ are 1.177, 1.268 and 1.190 with 95% HPD intervals of (0.824, 1.652), (0.849, 1.652) and (0.852, 1.660), for the nonrobust, partially robust and robust models, respectively. Note that the estimates arising from the three models are similar.

Based on the original data set, the inference arising from our robust model is the one that best reflects the behaviour of the bulk of the data, compared to the inferences arising from the non robust and partially robust models. Our robust model therefore succeeds in limiting the influence of outliers in order to obtain conclusions consistent with the majority of the observations.

Note that an outlier with an extreme x value, as $(x_{20}, y_{20}) = (696.4, 41.1)$, can be viewed as an observation with a fixed x value and an extreme y value (in this case, as an observation with a fixed x value of 696.4 and an extremely low y value of 41.1, compared to the trend emerging from the bulk of the data). This explains why our robust model produces robust inference in the presence of this type of outliers.

y_i	31.7	68.4	54.4	53.5	78.4	66.4	64.1	44.6	99.0	53.3
x_i	102.9	144.9	155.8	176.5	177.4	182.2	197.9	199.2	211.3	215.9
y_i	67.3	68.6	63.0	100.6	82.2	113.4	6.1	76.6	92.7	41.1
x_i	216.0	216.7	220.3	222.8	229.0	250.0	250.2	275.4	342.4	696.4

Table 2: Weekly expenditure on food (y_i) and weekly income (x_i) for household i in dollars, $i = 1, \dots, 20$

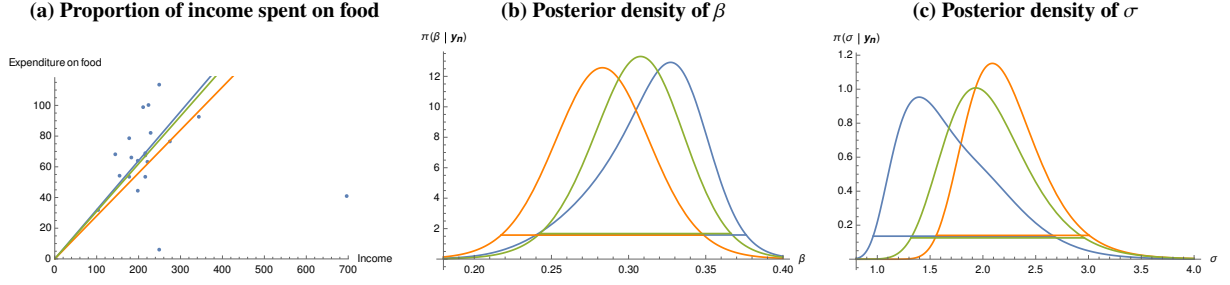


Figure 3: Expenditure on food as a function of the income with an estimation of the expenditure on food $\hat{\beta}x_i$ based on the posterior median, (b)-(c) Posterior densities of β and σ arising from the original data with 95% HPD intervals (horizontal lines); for each graph, the orange, green and blue lines are respectively related to the nonrobust, partially robust and robust models

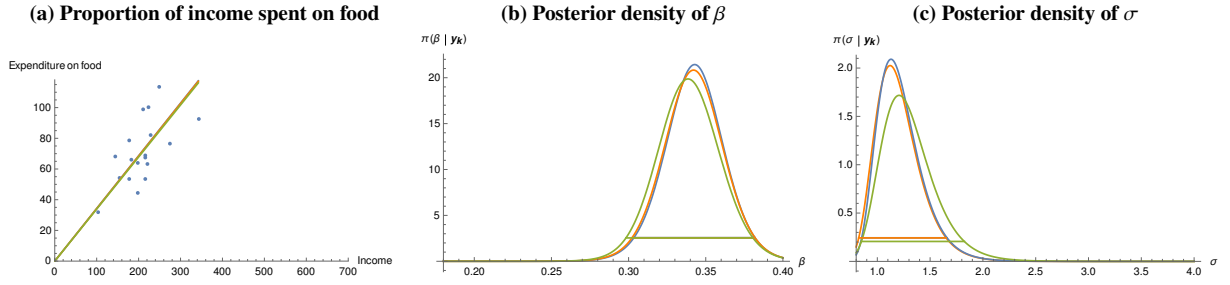


Figure 4: Expenditure on food as a function of the income with an estimation of the expenditure on food $\hat{\beta}x_i$ based on the posterior median, when the outliers are excluded, (b)-(c) Posterior densities of β and σ arising from the data set excluding the outliers with 95% HPD intervals (horizontal lines); for each graph, the orange, green and blue lines are respectively related to the nonrobust, partially robust and robust models

3.2. Simulation Study

We now compare the accuracy of the estimates arising from our robust model with those of the nonrobust and partially robust models, again in a context of finite population sampling. More precisely, the model $Y_i = \beta x_i + \epsilon_i$ with $\epsilon_i \mid \sigma \stackrel{\mathcal{D}}{\sim} 1/(\sigma x_i^{1/2})f(\epsilon_i/(\sigma x_i^{1/2}))$ and $x_i > 0$, $i = 1, \dots, n$, is used to analyse the data, where f is assumed to be a log-Pareto-tailed standard normal density with $\alpha = 1.96$ and $\phi = 4.08$ in our robust model, and it is compared with the same linear regression model, but where f is assumed to be a standard normal density in the nonrobust model, and where f is assumed to be a Student density with 10 degrees of freedom and a known scale parameter of 0.88 in the partially robust model. Note that the parameters of the distributions are the same as in Sect. 3.1.

For the simulation study, we set $n = 20$, $x_1, x_2, \dots, x_{20} = 1, 2, \dots, 20$, and $\pi(\beta, \sigma) \propto 1$. We simulate 50,000 data sets using values for β and σ arbitrarily set to 1 and 1.5, respectively, and we carry out this process for each of the three scenarios that we now describe. In the first one, f is a standard normal distribution; therefore, the probability to observe outliers is negligible. In the second scenario, f is a mixture of two normals where the first component is a standard normal

distribution and the second has a mean of 0 and a variance of 10^2 , with weights of 0.9 and 0.1, respectively. This last component can contaminate the data set by generating extreme values. In the third and last scenario, f is also a mixture of two normals, but the contamination is due to the second component's location. More precisely, the first component is again a standard normal, but the second has a mean of 10 and a variance of 1, with weights of 0.95 and 0.05, respectively.

For each simulated data set, we estimate β and σ using MAP estimation for the three models. Then, within each simulation scenario, we evaluate the performance of each model using sample mean square errors (MSE), based on the true values $\beta = 1$ and $\sigma = 1.5$. The results are presented in Tables 3 and 4.

We first observe that the three models have identical performances (to two decimal places) for both the estimation of β and σ , when there is no outliers (the 100% $N(0, 1)$ scenario). This was expected given that the three densities studied are very similar, especially on the interval $[-1.96, 1.96]$ where they all have 95% of their mass. They differ however in the thickness of their tails, and this feature plays a major role when the sample contains outliers, which is frequently the case for the two other scenarios. For the log-Pareto-tailed normal assumption, we observe that the MSE are minimally affected by the presence of outliers for both the estimation of β and σ , which confirms that the proposed approach provides whole robustness with respect to outliers. When it is assumed that the error has a Student distribution, outliers influence the estimation of σ significantly, while having a lesser effect on $\hat{\beta}$, which reflects the partial robustness of this approach. Finally, we observe that the presence of outlying observations has a major impact on the estimations when the traditional standard normal assumption is used.

Assumptions on f	Scenarios		
	$100\%N(0, 1)$	$90\%N(0, 1) + 10\%N(0, 10^2)$	$95\%N(0, 1) + 5\%N(10, 1)$
Standard normal	0.01	0.12	0.11
Student (10 d.f.)	0.01	0.03	0.04
Log-Pareto-tailed normal	0.01	0.02	0.03

Table 3: MSE of the estimators of β under the three scenarios and the three assumptions of f

Assumptions on f	Scenarios		
	$100\%N(0, 1)$	$90\%N(0, 1) + 10\%N(0, 10^2)$	$95\%N(0, 1) + 5\%N(10, 1)$
Standard normal	0.06	12.94	5.04
Student (10 d.f.)	0.06	4.07	2.02
Log-Pareto-tailed normal	0.06	0.75	0.48

Table 4: MSE of the estimators of σ under the three scenarios and the three assumptions of f

4. Conclusion

In this paper, we have provided a simple Bayesian approach to robustly estimate both parameters β and σ of a simple linear regression through the origin, in which the variance of the error

term can depend on the explanatory variable. It leads to robust estimators of finite population means and ratios. The approach is to replace the traditional normal assumption on the error term by a super heavy-tailed distribution assumption. In particular, we considered log-regularly varying distributions. Whole robustness is attained provided that both the negative and positive slope outliers are fewer than the nonoutlying observations, i.e. $m < k$ and $p < k$, as stated in Theorem 1. By whole robustness, we mean that the impact of outliers on the estimation gradually decrease to nothing as they converge towards plus or minus infinity.

The theoretical results have been illustrated in Sect. 3 through typical real-life situations in which ratio estimation is used, and a simulation study. All the analyses leading to robust inference have been done using the log-Pareto-tailed standard normal density given in (2). Our model has been compared with the nonrobust (with the normal assumption) and partially robust (with the Student distribution assumption) models. The conclusion is: our model performs as well as the nonrobust and the partially robust models in absence of outliers, in addition to being completely robust. Therefore, our recommendation is the following: assume that the error has the density given in (2) and obtain adequate results, regardless of whether there are outliers, by computing estimates as usual from the posterior distribution.

5. Proofs

Proposition 1 and Theorem 1 from our paper are proved in Sect. 5.1 and Sect. 5.2, respectively.

Beforehand, note that the assumptions on f imply that $f(z)$ and $|z|f(z)$ are bounded on the real line, with a limit of 0 in their tails as $|z| \rightarrow \infty$. As a result, we can define the constant $B > 0$ as follows:

$$B := \max \left\{ \sup_{z \in \mathbb{R}} f(z), \sup_{z \in \mathbb{R}} |z|f(z), \sup_{\beta \in \mathbb{R}, \sigma > 0} \min(\sigma, 1)\pi(\beta, \sigma) \right\}.$$

We also define the constant $\zeta \geq 1$ as follows:

$$\zeta := \max_i \left\{ \max \left\{ |x_i|, |x_i|^{-1} \right\} \right\}.$$

It results that for all $i \in \{1, \dots, n\}$ and for any $\varepsilon \in \mathbb{R}$, we have

$$\zeta^{-|\varepsilon|} \leq |x_i|^\varepsilon \leq \zeta^{|\varepsilon|}.$$

The monotonicity of the tails of $f(z)$ and $|z|f(z)$ implies that there exists a constant $M > 0$ such that

$$|y| \geq |z| \geq M \Rightarrow f(y) \leq f(z) \text{ and } |y|f(y) \leq |z|f(z). \quad (3)$$

5.1. Proof of Proposition 1

To prove that $\pi(\beta, \sigma \mid \mathbf{y}_n)$ is proper (the proof for $\pi(\beta, \sigma \mid \mathbf{y}_k)$ is omitted because it is similar), it suffices to show that the marginal $m(\mathbf{y}_n)$ is finite. Without loss of generality, we assume for convenience that $y_1/x_1 < \dots < y_n/x_n$. Note that we have strict inequalities because Y_1, \dots, Y_n are continuous random variables. Let the constant $\delta > 0$ be defined as

$$\delta = \zeta^{-1} \times \min_{i \in \{1, \dots, n-1\}} \{(y_{i+1}/x_{i+1} - y_i/x_i)/2\}.$$

We first show that the function is integrable on the area where the ratio $1/\sigma$ is bounded above. More precisely, we consider $\beta \in \mathbb{R}$ and $\delta(M\zeta^{|\theta|})^{-1} \leq \sigma < \infty$. Then, we show that the function is integrable on the area where the ratio $1/\sigma$ approaches infinity. We have

$$\begin{aligned} & \int_{\delta(M\zeta^{|\theta|})^{-1}}^{\infty} \int_{-\infty}^{\infty} \pi(\beta, \sigma) \prod_{i=1}^n \sigma^{-1}|x_i|^{-\theta} f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)) d\beta d\sigma \\ & \stackrel{a}{\leq} \max\left(\frac{1}{\sigma}, 1\right) B^n \zeta^{|\theta|(n-1)} \int_{\delta(M\zeta^{|\theta|})^{-1}}^{\infty} \frac{1}{\sigma^{n-1}} \int_{-\infty}^{\infty} \frac{1}{\sigma|x_1|^\theta} f\left(\frac{y_1 - \beta x_1}{\sigma|x_1|^\theta}\right) d\beta d\sigma \\ & \stackrel{b}{\leq} \max(\delta^{-1}M\zeta^{|\theta|}, 1) B^n \zeta^{|\theta|(n-1)} |x_1|^{-1} \int_{\delta(M\zeta^{|\theta|})^{-1}}^{\infty} \sigma^{-(n-1)} d\sigma \int_{-\infty}^{\infty} f(\beta') d\beta' \\ & \propto \int_{\delta(M\zeta^{|\theta|})^{-1}}^{\infty} \sigma^{-(n-1)} d\sigma \int_{-\infty}^{\infty} f(\beta') d\beta' \stackrel{c}{=} (\delta^{-1}M\zeta^{|\theta|})^{n-2} (n-2)^{-1} < \infty. \end{aligned}$$

In step *a*, we use $|x_i|^{-\theta} \leq \zeta^{|\theta|}$ for $i = 2, \dots, n$, and we bound $\min(\sigma, 1)\pi(\beta, \sigma)$ and each of $n-1$ densities f by B . In step *b*, we use the change of variable $\beta' = \sigma^{-1}|x_1|^{-\theta}(y_1 - \beta x_1)$. In step *c*, we use $n > 2$. Note that if instead, in step *a*, we bound $\sigma\pi(\beta, \sigma)$ by B , one can verify that the condition $n \geq 2$ is sufficient to bound above the integral.

We now show that the integral is finite on $\beta \in \mathbb{R}$ and $0 < \sigma < \delta(M\zeta^{|\theta|})^{-1}$. We have to carefully analyse the subareas where $y_i - \beta x_i$ is close to 0 in order to deal with the $0/0$ form of the ratios $(y_i - \beta x_i)/(\sigma|x_i|^\theta)$. In order to achieve this, we split the domain of β into n mutually exclusive areas as follows: $\mathbb{R} = \cup_{j=1}^n \{\beta : (y_{j-1}/x_{j-1} + y_j/x_j)/2 \leq \beta \leq (y_j/x_j + y_{j+1}/x_{j+1})/2\}$, where $y_0/x_0 := -\infty$ and $y_{n+1}/x_{n+1} := \infty$. We now consider $0 < \sigma < \delta(M\zeta^{|\theta|})^{-1}$ and $(y_{j-1}/x_{j-1} + y_j/x_j)/2 \leq \beta \leq (y_j/x_j + y_{j+1}/x_{j+1})/2$, $j \in \{1, \dots, n\}$.

$$\begin{aligned} & \pi(\beta, \sigma) \prod_{i=1}^n \sigma^{-1}|x_i|^{-\theta} f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)) \\ & \stackrel{a}{\leq} \sigma^{-1} B \max(1, \delta(M\zeta^{|\theta|})^{-1}) \prod_{i=1}^n \sigma^{-1}|x_i|^{-\theta} f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)) \\ & \propto \sigma^{-1} B \sigma^{-1}|x_j|^{-\theta} f(\sigma^{-1}|x_j|^{-\theta}(y_j - \beta x_j)) \prod_{i=1(i \neq j)}^n \sigma^{-1}|x_i|^{-\theta} f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)) \\ & \stackrel{b}{\leq} \sigma^{-1} B \sigma^{-1}|x_j|^{-\theta} f(\sigma^{-1}|x_j|^{-\theta}(y_j - \beta x_j)) \left[\sigma^{-1} \zeta^{|\theta|} f(\sigma^{-1} \zeta^{-|\theta|} \delta) \right]^{n-1} \\ & \stackrel{c}{\leq} B^{n-1} \zeta^{|\theta|(2n-3)} \delta^{-(n-2)} \sigma^{-1}|x_j|^{-\theta} f(\sigma^{-1}|x_j|^{-\theta}(y_j - \beta x_j)) \sigma^{-2} f(\sigma^{-1} \zeta^{-|\theta|} \delta) \\ & \propto \sigma^{-1}|x_j|^{-\theta} f(\sigma^{-1}|x_j|^{-\theta}(y_j - \beta x_j)) \sigma^{-2} \zeta^{-|\theta|} \delta f(\sigma^{-1} \zeta^{-|\theta|} \delta). \end{aligned}$$

In step *a*, we use $\pi(\beta, \sigma) \leq \max(\sigma^{-1}, 1)B = \sigma^{-1}B \max(1, \sigma) \leq \sigma^{-1}B \max(1, \delta(M\zeta^{|\theta|})^{-1})$. In step *b*, for $i \neq j$, we first note that

$$\begin{aligned} |y_i - \beta x_i| &= |x_i| |y_i/x_i - \beta| \geq \zeta^{-1} |y_i/x_i - \beta| \\ &\geq \zeta^{-1} \times \min \left\{ (y_j/x_j - y_{j-1}/x_{j-1})/2, (y_{j+1}/x_{j+1} - y_j/x_j)/2 \right\} \geq \delta, \end{aligned}$$

and then we use $f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)) \leq f(\sigma^{-1}\zeta^{-|\theta|}\delta)$ by the monotonicity of the tails of f since $\sigma^{-1}|x_i|^{-\theta}|y_i - \beta x_i| \geq \sigma^{-1}|x_i|^{-\theta}\delta \geq \sigma^{-1}\zeta^{-|\theta|}\delta \geq \delta^{-1}M\zeta^{|\theta|}\zeta^{-|\theta|}\delta = M$. Again for $i \neq j$, we use $|x_i|^{-\theta} \leq \zeta^{|\theta|}$. In step *c*, we bound $n - 2$ terms $\sigma^{-1}f(\sigma^{-1}\zeta^{-|\theta|}\delta)$ by $\zeta^{|\theta|}\delta^{-1}B$.

Finally, we have

$$\begin{aligned} &\int_0^{\delta(M\zeta^{|\theta|})^{-1}} \frac{\delta}{\sigma^2 \zeta^{|\theta|}} f\left(\frac{\delta}{\sigma \zeta^{|\theta|}}\right) \int_{(y_{j-1}/x_{j-1} + y_j/x_j)/2}^{(y_j/x_j + y_{j+1}/x_{j+1})/2} \frac{1}{\sigma |x_j|^\theta} f\left(\frac{y_j - \beta x_j}{\sigma |x_j|^\theta}\right) d\beta d\sigma \\ &\leq |x_j|^{-1} \int_0^\infty f(\sigma') d\sigma' \int_{-\infty}^\infty f(\beta') d\beta' \leq |x_j|^{-1} \leq \zeta < \infty, \end{aligned}$$

where we use the change of variables $\sigma' = \sigma^{-1}\zeta^{-|\theta|}\delta$ and $\beta' = \sigma^{-1}|x_j|^{-\theta}(y_j - \beta x_j)$. Note that we do not need to assume that f is a log-regularly varying distribution to obtain the result.

5.2. Proof of Theorem 1

Consider the model and the context described in Sect. 2.1. We assume that $zf(z) \in L_\rho(\infty)$ and $k > \max(m, p)$. In addition, we assume that $m + p \geq 1$, i.e. that there is at least one outlier, otherwise the proof would be trivial. Two lemmas are first given and the proofs of results (a) to (e) follows. The proofs of these two lemmas can be found in [Desgagné \(2015\)](#).

Lemma 1. $\forall \lambda \geq 0, \forall \tau \geq 1$, there exists a constant $D(\lambda, \tau) \geq 1$ such that $z \in \mathbb{R}$ and $(\mu, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau] \Rightarrow$

$$1/D(\lambda, \tau) \leq (1/\sigma)f((z - \mu)/\sigma)/f(z) \leq D(\lambda, \tau).$$

Note that Lemma 1 is a corollary of Proposition 4 of [Desgagné \(2015\)](#).

Lemma 2. There exists a constant $C > 0$ such that

$$|z| \geq 2M \Rightarrow \sup_{\mu \in \mathbb{R}} \frac{f(\mu)f(z - \mu)}{f(z)} \leq C,$$

where M is given in equation (3).

Proof of Result (a). We first observe that

$$\begin{aligned}
& \frac{m(\mathbf{y}_n)}{m(\mathbf{y}_k) \prod_{i=1}^n [f(y_i)]^{m_i+p_i}} \\
&= \frac{m(\mathbf{y}_n)}{m(\mathbf{y}_k) \prod_{i=1}^n [f(y_i)]^{m_i+p_i}} \int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta, \sigma \mid \mathbf{y}_n) d\sigma d\beta \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \frac{\pi(\beta, \sigma) \prod_{i=1}^n \left[\sigma^{-1} |x_i|^{-\theta} f\left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)\right) \right]^{k_i+m_i+p_i}}{m(\mathbf{y}_k) \prod_{i=1}^n [f(y_i)]^{m_i+p_i}} d\sigma d\beta \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f\left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)\right)}{f(y_i)} \right]^{m_i+p_i} d\sigma d\beta.
\end{aligned}$$

We show that the last integral converges to 1 as $\omega \rightarrow \infty$ to prove result (a). If we use Lebesgue's dominated convergence theorem to interchange the limit $\omega \rightarrow \infty$ and the integral, we have

$$\begin{aligned}
& \lim_{\omega \rightarrow \infty} \int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f\left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)\right)}{f(y_i)} \right]^{m_i+p_i} d\sigma d\beta \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \lim_{\omega \rightarrow \infty} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f\left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)\right)}{f(y_i)} \right]^{m_i+p_i} d\sigma d\beta \\
&= \int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta, \sigma \mid \mathbf{y}_k) d\sigma d\beta = 1,
\end{aligned}$$

using Proposition 4 of [Desgagné \(2015\)](#) in the second equality, since x_1, \dots, x_n and θ are fixed, and then Proposition 1. Note that pointwise convergence is sufficient, for any value of $\beta \in \mathbb{R}$ and $\sigma > 0$, once the limit is inside the integral. However, in order to use Lebesgue's dominated convergence theorem, we need to show that the integrand is bounded, for any value of $\omega \geq y$, by an integrable function of β and σ that does not depend on ω . The constant y can be chosen as large as we want, and minimum values for y will be given throughout the proof. In order to bound the integrand, we divide the domain of integration into four quadrants delineated by the axes $\beta = 0$ and $\sigma = 1$. The proofs are given only for the two quadrants where $\beta \geq 0$ because the proofs for $\beta < 0$ are similar. The strategy is again to separately analyse the area where the ratio $1/\sigma$ approaches infinity.

We assumed that y_i can be written as $y_i = a_i + b_i \omega$, where $\omega \rightarrow \infty$, a_i and b_i are constants such that $a_i \in \mathbb{R}$ and $b_i \neq 0$ if y_i is an outlier. Therefore, the ranking of the elements in the set $\{|y_i| : m_i + p_i = 1\}$ is primarily determined by the values $|b_1|, \dots, |b_n|$ and we can choose the constant y larger than a certain threshold such that this ranking remains unchanged for all $\omega \geq y$. Without loss of generality, we assume for convenience that

$$\omega = \min_{\{i : m_i+p_i=1\}} |y_i| \quad \text{and consequently} \quad \min_{\{i : m_i+p_i=1\}} |b_i| = 1,$$

and we also assume that y_1 is a nonoutlier (therefore $k_1 = 1$). We now bound above the integrand on the first quadrant.

Quadrant 1: Consider $0 \leq \beta < \infty$ and $1 \leq \sigma < \infty$. We have

$$\begin{aligned}
& \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} \\
& \propto \frac{\pi(\beta, \sigma)}{\sigma^n} \prod_{i=1}^n \frac{|x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{[f(y_i)]^{m_i + p_i}} \\
& \leq \frac{B}{\sigma^n} \prod_{i=1}^n \frac{D(|a_i|, \zeta^{|\theta|}) f((b_i \omega - \beta x_i)/\sigma)}{[f(y_i)]^{m_i + p_i}} \\
& \leq \frac{1}{[f(\omega)]^{m+p}} \frac{B}{\sigma^n} \prod_{i=1}^n D(|a_i|, \zeta^{|\theta|}) f((b_i \omega - \beta x_i)/\sigma) [|b_i| D(|a_i|, |b_i|)]^{m_i + p_i} \\
& \propto \frac{1}{[f(\omega)]^{m+p}} \frac{1}{\sigma^n} \prod_{i=1}^n f((b_i \omega - \beta x_i)/\sigma) \\
& \stackrel{c}{=} \frac{1}{[f(\omega)]^{m+p}} \frac{1}{\sigma^n} \prod_{i=1}^n [f(\beta x_i/\sigma)]^{k_i} [f((b_i \omega - \beta x_i)/\sigma)]^{m_i + p_i} \\
& \stackrel{d}{=} \frac{(1/\sigma) f(\beta x_1/\sigma)}{\sigma^{k-3/2}} \left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i} [f((b_i \omega - \beta x_i)/\sigma)]^{m_i + p_i}.
\end{aligned}$$

In step *a*, we use $y_i = a_i + b_i \omega$ and Lemma 1 to obtain

$$\frac{1}{|x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right) = \frac{1}{|x_i|^\theta} f\left(\frac{(b_i \omega - \beta x_i)/\sigma + a_i/\sigma}{|x_i|^\theta}\right) \leq D(|a_i|, \zeta^{|\theta|}) f((b_i \omega - \beta x_i)/\sigma)$$

because $|a_i/\sigma| \leq |a_i|$ and $\zeta^{-|\theta|} \leq |x_i|^\theta \leq \zeta^{|\theta|}$, for all i . We also use $\pi(\beta, \sigma) \leq \max(\sigma^{-1}, 1)B = B$. In step *b*, we again use Lemma 1 to obtain $f(\omega)/f(y_i) = f((y_i - a_i)/b_i)/f(y_i) \leq |b_i| D(|a_i|, |b_i|)$. In step *c*, we set $b_i = 0$ if $k_i = 1$ and we use the symmetry of f to obtain $f(-\beta x_i/\sigma) = f(\beta x_i/\sigma)$. In step *d*, we use the assumption $k_1 = 1$, which implies that $m_1 = p_1 = 0$.

Now it suffices to demonstrate that

$$\left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i} [f((b_i \omega - \beta x_i)/\sigma)]^{m_i + p_i} \quad (4)$$

is bounded by a constant that does not depend on ω, β and σ since $(1/\sigma)^{k-3/2} (1/\sigma) f(\beta x_1/\sigma)$ is an integrable function on quadrant 1. Indeed, since $k > 2$, we have

$$\int_1^\infty (1/\sigma)^{k-3/2} \int_0^\infty (1/\sigma) f(\beta x_1/\sigma) d\beta d\sigma \leq |x_1|^{-1} \int_1^\infty (1/\sigma)^{k-3/2} d\sigma = \frac{|x_1|^{-1}}{k-5/2} \leq 2\zeta.$$

Note that if instead, in step *a*, we bound $\pi(\beta, \sigma)$ by $\sigma^{-1}B$, one can verify that the condition $k \geq 2$ is sufficient to bound above the integral.

In order to bound above the function in (4), we separately analyse the three following cases: ω/σ is large, ω/σ is either large or bounded, and ω/σ is bounded. More precisely, we split quadrant 1 with respect to σ into three parts: $1 \leq \sigma < \omega^{1/2}$, $\omega^{1/2} \leq \sigma < \omega/(2M)$ and $\omega/(2M) \leq \sigma < \infty$, where M is defined in equation (3). Note that this is well defined if $y > \max(1, (2M)^2)$ since $\omega \geq y$.

First, we consider $0 \leq \beta < \infty$ and $\omega/(2M) \leq \sigma < \infty$. We have,

$$\begin{aligned} \left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i} [f((b_i\omega - \beta x_i)/\sigma)]^{m_i+p_i} &\stackrel{a}{\leq} B^{n-1} \left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \\ &\stackrel{b}{\leq} B^{n-1} (2M)^{m+p+1/2} \frac{(1/\omega)^{1/2}}{[\omega f(\omega)]^{m+p}} \stackrel{c}{\leq} B^{n-1} (2M)^{m+p+1/2} \frac{(1/\omega)^{1/2}}{(\log \omega)^{-(\rho+1)(m+p)}} \\ &\stackrel{d}{\leq} B^{n-1} (2M)^{m+p+1/2} [2(\rho+1)(m+p)/e]^{(\rho+1)(m+p)} < \infty. \end{aligned}$$

In step *a*, we use $f \leq B$. In step *b*, we use $\omega/\sigma \leq 2M$ and $(1/\sigma) \leq (2M)/\omega$. In step *c*, we use $\omega f(\omega) > (\log \omega)^{-\rho-1}$ if $\omega \geq y \geq A(1)$, where $A(1)$ comes from Proposition 2 of [Desgagné \(2015\)](#). For step *d*, it is purely algebraic to show that the maximum of $(\log \omega)^\xi / \omega^{1/2}$ is $(2\xi/e)^\xi$ for $\omega > 1$ and $\xi > 0$, where $\xi = (\rho+1)(m+p)$ in our situation.

Now, consider the two other parts combined (we will split them in the next step), that is $0 \leq \beta < \infty$ and $1 \leq \sigma \leq \omega/(2M)$. We have,

$$\begin{aligned} &\left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i} [f((b_i\omega - \beta x_i)/\sigma)]^{m_i+p_i} \\ &\stackrel{a}{\leq} \left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i} [f(b_i\omega/\sigma)]^{m_i} [f((b_i\omega - \beta x_i)/\sigma)]^{p_i} \\ &= \left[\frac{\omega/\sigma}{\omega f(\omega)} \right]^{m+p} \frac{1}{\sigma^{1/2}} \prod_{i=2}^n \left[f\left(\frac{\beta x_i}{\sigma}\right) \right]^{k_i-p_i} \left[f\left(\frac{b_i\omega}{\sigma}\right) \right]^{m_i+p_i} \left[\frac{f\left(\frac{b_i\omega}{\sigma} - \frac{\beta x_i}{\sigma}\right) f\left(\frac{\beta x_i}{\sigma}\right)}{f\left(\frac{b_i\omega}{\sigma}\right)} \right]^{p_i} \\ &\stackrel{b}{\leq} C^p \frac{1}{\sigma^{1/2}} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i-p_i} \left[\frac{(\omega/\sigma)f(b_i\omega/\sigma)}{\omega f(\omega)} \right]^{m_i+p_i} \\ &\stackrel{c}{\leq} C^p \frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p} \prod_{i=2}^n [f(\beta x_i/\sigma)]^{k_i-p_i} \\ &\stackrel{d}{\leq} C^p \frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p} [f(\beta/\sigma)]^{k-1-p} [\zeta D(0, \zeta)]^{k-1+p} \\ &\stackrel{e}{\leq} C^p \frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p} B^{k-1-p} [\zeta D(0, \zeta)]^{k-1+p} \propto \frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p}. \end{aligned}$$

In step *a*, we use $f((b_i\omega - \beta x_i)/\sigma) \leq f(b_i\omega/\sigma)$ if $m_i = 1$ (in this case $x_i > 0, b_i < 0$ or $x_i < 0, b_i > 0$) by the monotonicity of the tails of f since $|b_i\omega - \beta x_i|/\sigma = (|b_i|\omega + \beta|x_i|)/\sigma \geq |b_i|\omega/\sigma$

$\geq |b_i|(2M) \geq 2M \geq M$. In step *b*, we use Lemma 2 since $|b_i|\omega/\sigma \geq |b_i|(2M) \geq 2M$. In step *c*, we use $f(b_i\omega/\sigma) \leq f(\omega/\sigma)$ by the monotonicity of the tails of f since $|b_i|\omega/\sigma \geq \omega/\sigma \geq 2M \geq M$. In step *d*, we use Lemma 1 to obtain $f(\beta|x_i|/\sigma) \leq |x_i|^{-1}D(0, \zeta)f(\beta/\sigma) \leq \zeta D(0, \zeta)f(\beta/\sigma)$, and similarly $1/f(\beta|x_i|/\sigma) \leq \zeta D(0, \zeta)/f(\beta/\sigma)$. In step *e*, we use $[f(\beta/\sigma)]^{k-1-p} \leq B^{k-1-p}$ since $k-1 \geq p$ (by assumption $k > \max(m, p) \Rightarrow k > p$).

Now, we consider $0 \leq \beta < \infty$ and $\omega^{1/2} \leq \sigma \leq \omega/(2M)$. We have,

$$\begin{aligned} \frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p} &\stackrel{a}{\leq} B^{m+p} \frac{(1/\omega)^{1/4}}{[\omega f(\omega)]^{m+p}} \stackrel{b}{\leq} B^{m+p} \frac{(1/\omega)^{1/4}}{(\log \omega)^{-(\rho+1)(m+p)}} \\ &\stackrel{c}{\leq} B^{m+p} [4(\rho+1)(m+p)/e]^{(\rho+1)(m+p)} < \infty. \end{aligned}$$

In step *a*, we use $(\omega/\sigma)f(\omega/\sigma) \leq B$ and $(1/\sigma)^{1/2} \leq (1/\omega)^{1/4}$. In step *b*, we use $\omega f(\omega) > (\log \omega)^{-\rho-1}$ if $\omega \geq y \geq A(1)$, where $A(1)$ comes from Proposition 2 of [Desgagné \(2015\)](#). In step *c*, it is purely algebraic to show that the maximum of $(\log \omega)^\xi / \omega^{1/4}$ is $(4\xi/e)^\xi$ for $\omega > 1$ and $\xi > 0$, where $\xi = (\rho+1)(m+p)$ in our situation.

Finally, we consider $0 \leq \beta < \infty$ and $1 \leq \sigma \leq \omega^{1/2}$. We have,

$$\frac{1}{\sigma^{1/2}} \left[\frac{(\omega/\sigma)f(\omega/\sigma)}{\omega f(\omega)} \right]^{m+p} \stackrel{a}{\leq} \left[\frac{\omega^{1/2}f(\omega^{1/2})}{\omega f(\omega)} \right]^{m+p} \stackrel{b}{\leq} 2^{(\rho+1)(m+p)} < \infty.$$

In step *a*, we use $1/\sigma \leq 1$ and we use $(\omega/\sigma)f(\omega/\sigma) \leq \omega^{1/2}f(\omega^{1/2})$ by the monotonicity of the tails of $|z|f(z)$ since $\omega/\sigma \geq \omega^{1/2} \geq y^{1/2} \geq M$ if $y \geq M^2$. In step *b*, we use $\omega^{1/2}f(\omega^{1/2})/(\omega f(\omega)) \leq 2(1/2)^{-\rho} = 2^{\rho+1}$ if $\omega \geq y \geq A(1, 2)$, where $A(1, 2)$ comes from the definition of a log-regularly varying function (see Definition 1 of [Desgagné \(2015\)](#)).

Quadrant 2: Consider $-\infty < \beta < 0$ and $1 \leq \sigma < \infty$. The proof for quadrant 2 is similar to that of quadrant 1. The condition $k > p$ is replaced by $k > m$. Note that $k > \max(m, p)$ is assumed in Theorem 1.

Quadrant 3: Consider $-\infty < \beta < 0$ and $0 < \sigma < 1$. The proof for quadrant 3 is similar to that of quadrant 4, given below. The condition $k > p$ is replaced by $k > m$. Note that $k > \max(m, p)$ is assumed in Theorem 1.

Quadrant 4: Consider $0 \leq \beta < \infty$ and $0 < \sigma < 1$. We actually need to show that

$$\begin{aligned} \lim_{\omega \rightarrow \infty} \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1}|x_i|^{-\theta} f(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i))}{f(y_i)} \right]^{m_i+p_i} d\sigma d\beta \\ = \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) d\sigma d\beta. \end{aligned}$$

For quadrant 4, we proceed in a slightly different manner than for quadrant 1. We begin by separating the first integral into two parts as follows:

$$\begin{aligned}
& \lim_{\omega \rightarrow \infty} \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} d\sigma d\beta \\
&= \lim_{\omega \rightarrow \infty} \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\frac{1}{\sigma |x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right)}{f(y_i)} \right]^{m_i + p_i} \mathbb{1}_{[0, \zeta^{-1}\omega/2]}(\beta) d\sigma d\beta \\
&+ \lim_{\omega \rightarrow \infty} \int_{\zeta^{-1}\omega/2}^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} d\sigma d\beta,
\end{aligned}$$

where the indicator function $\mathbb{1}_A(\beta)$ is equal to 1 if $\beta \in A$, and equal to 0 otherwise. We show that the first part is equal to the integral $\int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) d\sigma d\beta$ and the second part is equal to 0.

For the first part, we again use Lebesgue's dominated convergence theorem in order to interchange the limit $\omega \rightarrow \infty$ and the integral. We have

$$\begin{aligned}
& \lim_{\omega \rightarrow \infty} \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\frac{1}{\sigma |x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right)}{f(y_i)} \right]^{m_i + p_i} \mathbb{1}_{[0, \zeta^{-1}\omega/2]}(\beta) d\sigma d\beta \\
&= \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \lim_{\omega \rightarrow \infty} \prod_{i=1}^n \left[\frac{\frac{1}{\sigma |x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma |x_i|^\theta}\right)}{f(y_i)} \right]^{m_i + p_i} \mathbb{1}_{[0, \zeta^{-1}\omega/2]}(\beta) d\sigma d\beta \\
&= \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \times 1 \times \mathbb{1}_{[0, \infty)}(\beta) d\sigma d\beta = \int_0^\infty \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) d\sigma d\beta,
\end{aligned}$$

using Proposition 4 of [Desgagné \(2015\)](#) in the second equality since x_1, \dots, x_n and θ are fixed. Note that pointwise convergence is sufficient, for any value of $\beta \in \mathbb{R}$ and $\sigma > 0$, once the limit is inside the integral. We now demonstrate that the integrand is bounded, for any value of $\omega \geq y$, by an integrable function of β and σ that does not depend on ω .

Consider $0 \leq \beta \leq \zeta^{-1}\omega/2$ (the integrand is equal to 0 if $\zeta^{-1}\omega/2 < \beta < \infty$) and $0 < \sigma < 1$. We have

$$\begin{aligned}
& \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} \mathbb{1}_{[0, \zeta^{-1}\omega/2]}(\beta) \\
&\stackrel{a}{\leq} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\zeta^{-|\theta|} f(\zeta^{-|\theta|} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} \\
&\stackrel{b}{\leq} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\zeta^{-|\theta|} f(\zeta^{-|\theta|} \omega/2)}{f(y_i)} \right]^{m_i + p_i} \\
&\stackrel{c}{\leq} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n [2|b_i|D(|a_i|, 2|b_i|\zeta^{|\theta|})]^{m_i + p_i},
\end{aligned}$$

and $\pi(\beta, \sigma \mid \mathbf{y}_k)$ is an integrable function. In step *a*, we use the equality $\mathbb{1}_{[0, \zeta^{-1}\omega/2]}(\beta) = 1$. We also use

$$\sigma^{-1}|x_i|^{-\theta}|y_i - \beta x_i|f\left(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)\right) \leq \zeta^{-|\theta|}|y_i - \beta x_i|f\left(\zeta^{-|\theta|}(y_i - \beta x_i)\right)$$

by the monotonicity of the tails of $|z|f(z)$ and therefore we obtain

$$\sigma^{-1}|x_i|^{-\theta}f\left(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)\right) \leq \zeta^{-|\theta|}f\left(\zeta^{-|\theta|}(y_i - \beta x_i)\right),$$

and in step *b*, we use

$$f\left(\zeta^{-|\theta|}(y_i - \beta x_i)\right) \leq f(\zeta^{-|\theta|}\omega/2)$$

by the monotonicity of the tails of $f(z)$. Indeed, if $m_i = 1$ (in this case $x_i > 0, b_i < 0$ or $x_i < 0, b_i > 0$), we have $\sigma^{-1}|x_i|^{-\theta}|y_i - \beta x_i| \geq |x_i|^{-\theta}|y_i - \beta x_i| \geq \zeta^{-|\theta|}|y_i - \beta x_i| = \zeta^{-|\theta|}(|y_i| + \beta|x_i|) \geq \zeta^{-|\theta|}|y_i| \geq \zeta^{-|\theta|}\omega \geq \zeta^{-|\theta|}\omega/2 \geq \zeta^{-|\theta|}y/2 \geq M$, if we choose $y \geq 2\zeta^{|\theta|}M$. And, if $p_i = 1$, we have $\sigma^{-1}|x_i|^{-\theta}|y_i - \beta x_i| \geq \zeta^{-|\theta|}|y_i - \beta x_i| \geq \zeta^{-|\theta|}(|y_i| - \beta|x_i|) \geq \zeta^{-|\theta|}(\omega - (\zeta^{-1}\omega/2)\zeta) = \zeta^{-|\theta|}\omega/2 \geq \zeta^{-|\theta|}y/2 \geq M$. Note that $0 \leq \beta \leq \zeta^{-1}\omega/2$ is used only for the case $p_i = 1$ ($\beta \geq 0$ is sufficient for the case $m_i = 1$). In step *c*, we use Lemma 1 to obtain

$$\frac{f(\zeta^{-|\theta|}\omega/2)}{f(y_i)} = \frac{f((y_i - a_i)/(2b_i\zeta^{|\theta|}))}{f(y_i)} \leq 2|b_i|\zeta^{|\theta|}D(|a_i|, 2|b_i|\zeta^{|\theta|}).$$

We now prove that

$$\lim_{\omega \rightarrow \infty} \int_{\zeta^{-1}\omega/2}^{\infty} \int_0^1 \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1}|x_i|^{-\theta}f\left(\sigma^{-1}|x_i|^{-\theta}(y_i - \beta x_i)\right)}{f(y_i)} \right]^{m_i + p_i} d\sigma d\beta = 0.$$

We first bound above the integrand and then we prove that the integral of the upper bound converges towards 0 as $\omega \rightarrow \infty$.

Consider $\zeta^{-1}\omega/2 < \beta < \infty$ and $0 < \sigma < 1$. We have

$$\begin{aligned}
& \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(y_i)} \right]^{m_i + p_i} \\
& \stackrel{a}{\leq} \pi(\beta, \sigma \mid \mathbf{y}_k) \prod_{i=1}^n [2|b_i|D(|a_i|, 2|b_i|\zeta^{|\theta|})]^{m_i} \left[\frac{|b_i|D(|a_i|, |b_i|) \frac{1}{\sigma|x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma|x_i|^\theta}\right)}{f(\omega)} \right]^{p_i} \\
& \propto \pi(\beta, \sigma) \prod_{i=1}^n \left[\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (a_i - \beta x_i)) \right]^{k_i} \left[\frac{\frac{1}{\sigma|x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma|x_i|^\theta}\right)}{f(\omega)} \right]^{p_i} \\
& \stackrel{b}{\leq} \sigma^{-1} B \left[4\zeta^{2|\theta|+2} D(0, 4\zeta^{2+|\theta|}) (1/\sigma) f(\omega/\sigma) \right]^k \prod_{i=1}^n \left[\frac{\frac{1}{\sigma|x_i|^\theta} f\left(\frac{y_i - \beta x_i}{\sigma|x_i|^\theta}\right)}{f(\omega)} \right]^{p_i} \\
& \propto \sigma^{-1} \left[\sigma^{-1} f(\sigma^{-1} \omega) \right]^k \prod_{i=1}^n \left[\frac{\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i))}{f(\omega)} \right]^{p_i} \\
& \stackrel{c}{\leq} \sigma^{-1} \left[\sigma^{-1} f(\sigma^{-1} \omega) \right]^{k-p} \prod_{i=1}^n \left[\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)) \right]^{p_i} \\
& \stackrel{d}{=} \sigma^{-1} \left[\sigma^{-1} f(\sigma^{-1} \omega) \right]^{k-p} \prod_{i=1}^p \sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)).
\end{aligned}$$

In step *a*, for the case $m_i = 1$, we use the inequality $\sigma^{-1} |x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i)) / f(y_i) \leq 2|b_i|D(|a_i|, 2|b_i|\zeta^{|\theta|})$ by the same arguments used for the first part (steps *a* to *c*). Note that we still have $\beta \geq 0$ ($0 \leq \beta \leq \zeta^{-1}\omega/2$ was used only for the case $p_i = 1$). For the case $p_i = 1$, we use Lemma 1 to obtain $f(\omega)/f(y_i) = f((y_i - a_i)/b_i)/f(y_i) \leq |b_i|D(|a_i|, |b_i|)$. In step *b*, we use $\pi(\beta, \sigma) \leq \max(\sigma^{-1}, 1)B = \sigma^{-1} \max(1, \sigma)B = \sigma^{-1}B$. For the case $k_i = 1$, we use the monotonicity of the tails of f to obtain

$$|x_i|^{-\theta} f(\sigma^{-1} |x_i|^{-\theta} (a_i - \beta x_i)) \leq \zeta^{|\theta|} f(\sigma^{-1} \zeta^{-(2+|\theta|)} \omega/4) \leq 4\zeta^{2|\theta|+2} D(0, 4\zeta^{2+|\theta|}) f(\omega/\sigma)$$

because, if we define the constant $a_{(k)} := \max_{\{i: k_i=1\}} |a_i|$ with $\omega \geq y \geq 4\zeta^2 a_{(k)}$, we have $\sigma^{-1} |x_i|^{-\theta} |a_i - \beta x_i| \geq \sigma^{-1} |x_i|^{-\theta} (\beta |x_i| - |a_i|) \geq \sigma^{-1} \zeta^{-|\theta|} ((\zeta^{-1}\omega/2)\zeta^{-1} - a_{(k)}) \geq \sigma^{-1} \zeta^{-|\theta|} (\zeta^{-2}\omega/2 - \zeta^{-2}\omega/4) = \sigma^{-1} \zeta^{-(2+|\theta|)} \omega/4 \geq \zeta^{-(2+|\theta|)} \omega/4 \geq \zeta^{-(2+|\theta|)} y/4 \geq M$ if we choose $y \geq 4\zeta^{2+|\theta|}M$. We use Lemma 1 in the second inequality. In step *c*, we use the monotonicity of the tails of $|z|f(z)$ to obtain $\sigma^{-1}\omega f(\sigma^{-1}\omega) \leq \omega f(\omega)$ because $\sigma^{-1}\omega \geq \omega \geq y \geq M$ if we choose $y \geq M$. In step *d*, we assume for convenience and without loss of generality that $\{i : p_i = 1\} = \{1, \dots, p\}$, and we consider this assumption for the rest of the proof.

As in the proof of Proposition 1, we now split the real line (which includes $\zeta^{-1}\omega/2 \leq \beta < \infty$) into p mutually disjoint intervals given by $(y_{j-1}/x_{j-1} + y_j/x_j)/2 \leq \beta \leq (y_j/x_j + y_{j+1}/x_{j+1})/2$, for $j = 1, \dots, p$, where we define $y_0/x_0 := -\infty$ and $y_{p+1}/x_{p+1} := \infty$. We also define the constant $\delta > 0$ as follows:

$$\delta = \zeta^{-1} \times \min_{i \in \{1, \dots, p-1\}} \{(y_{i+1}/x_{i+1} - y_i/x_i)/2\}.$$

Consider $(y_{j-1}/x_{j-1} + y_j/x_j)/2 \leq \beta \leq (y_j/x_j + y_{j+1}/x_{j+1})/2$, for $j \in \{1, \dots, p\}$, and $0 < \sigma < 1$. Thus,

$$\begin{aligned} & \sigma^{-1} \left[\sigma^{-1} f(\sigma^{-1} \omega) \right]^{k-p} \prod_{i=1}^p \sigma^{-1} |x_i|^{-\theta} f \left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i) \right) \\ & \stackrel{a}{\leq} (\delta^{-1} B)^{p-1} \sigma^{-1} \left[\sigma^{-1} f(\sigma^{-1} \omega) \right]^{k-p} \sigma^{-1} |x_j|^{-\theta} f \left(\sigma^{-1} |x_j|^{-\theta} (y_j - \beta x_j) \right) \\ & \stackrel{b}{\leq} (\delta^{-1} B)^{p-1} B^{k-p-1} \omega^{-(k-p)} \sigma^{-2} \omega f(\sigma^{-1} \omega) \times \sigma^{-1} |x_j|^{-\theta} f \left(\sigma^{-1} |x_j|^{-\theta} (y_j - \beta x_j) \right). \end{aligned}$$

In step *a*, we use, for $i \neq j$, $\sigma^{-1} |x_i|^{-\theta} f \left(\sigma^{-1} |x_i|^{-\theta} (y_i - \beta x_i) \right) \leq |y_i - \beta x_i|^{-1} B \leq \delta^{-1} B$, where we bound $|z|f(z)$ by B and we use $|y_i - \beta x_i| \geq \delta$ since

$$\begin{aligned} |y_i - \beta x_i| &= |x_i| |y_i/x_i - \beta| \geq \zeta^{-1} |y_i/x_i - \beta| \\ &\geq \zeta^{-1} \times \min \left\{ (y_j/x_j - y_{j-1}/x_{j-1})/2, (y_{j+1}/x_{j+1} - y_j/x_j)/2 \right\} \geq \delta. \end{aligned}$$

In step *b*, we use $\sigma^{-1} \omega f(\sigma^{-1} \omega) \leq B$ for $k - p - 1$ terms (by assumption $k > \max(m, p) \Rightarrow k > p$).

Finally, we have

$$\begin{aligned} & \omega^{-(k-p)} \int_0^1 \sigma^{-2} \omega f(\sigma^{-1} \omega) \int_{(y_{j-1}/x_{j-1} + y_j/x_j)/2}^{(y_j/x_j + y_{j+1}/x_{j+1})/2} \sigma^{-1} |x_j|^{-\theta} f \left(\sigma^{-1} |x_j|^{-\theta} (y_j - \beta x_j) \right) d\beta d\sigma \\ & \leq \omega^{-(k-p)} \int_0^\infty \sigma^{-2} \omega f(\sigma^{-1} \omega) \int_{-\infty}^\infty \sigma^{-1} |x_j|^{-\theta} f \left(\sigma^{-1} |x_j|^{-\theta} (y_j - \beta x_j) \right) d\beta d\sigma \\ & \stackrel{a}{=} |x_j|^{-1} \omega^{-(k-p)} \int_0^\infty f(\sigma') d\sigma' \int_{-\infty}^\infty f(\beta') d\beta' \leq \zeta \omega^{-(k-p)} \xrightarrow{b} 0 \text{ as } \omega \rightarrow \infty. \end{aligned}$$

In step *a*, we use the change of variables $\sigma' = \sigma^{-1} \omega$ and $\beta' = \sigma^{-1} |x_j|^{-\theta} (y_j - \beta x_j)$. In step *b*, we use $k > p$. \square

Proof of Result (b). Consider (β, σ) such that $\pi(\beta, \sigma) > 0$ (the proof for the case (β, σ) such that $\pi(\beta, \sigma) = 0$ is trivial). We have, as $\omega \rightarrow \infty$,

$$\begin{aligned} \frac{\pi(\beta, \sigma \mid \mathbf{y}_n)}{\pi(\beta, \sigma \mid \mathbf{y}_k)} &= \frac{m(\mathbf{y}_k)}{m(\mathbf{y}_n)} \times \frac{\pi(\beta, \sigma) \prod_{i=1}^n (\sigma |x_i|^\theta)^{-1} f \left((\sigma |x_i|^\theta)^{-1} (y_i - \beta x_i) \right)}{\pi(\beta, \sigma) \prod_{i=1}^n [(\sigma |x_i|^\theta)^{-1} f \left((\sigma |x_i|^\theta)^{-1} (y_i - \beta x_i) \right)]^{k_i}} \\ &= \frac{m(\mathbf{y}_k)}{m(\mathbf{y}_n)} \prod_{i=1}^n \left[(\sigma |x_i|^\theta)^{-1} f \left((\sigma |x_i|^\theta)^{-1} (y_i - \beta x_i) \right) \right]^{m_i + p_i} \\ &= \frac{m(\mathbf{y}_k) \prod_{i=1}^n [f(y_i)]^{m_i + p_i}}{m(\mathbf{y}_n)} \prod_{i=1}^n \left[\frac{(\sigma |x_i|^\theta)^{-1} f \left((\sigma |x_i|^\theta)^{-1} (y_i - \beta x_i) \right)}{f(y_i)} \right]^{m_i + p_i} \rightarrow 1. \end{aligned}$$

The first ratio in the last equality does not depend on β and σ , and converges towards 1 as $\omega \rightarrow \infty$ using result (a). The second part also converges to 1 uniformly in any set $(\beta, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$ using Proposition 4 of [Desgagné \(2015\)](#) since x_1, \dots, x_n and θ are fixed. Furthermore,

since f and $\sigma\pi(\beta, \sigma)$ are bounded, and $x_i \neq 0$ for all i , $\pi(\beta, \sigma \mid \mathbf{y}_k)$ is also bounded on any set $(\beta, \sigma) \in [-\lambda, \lambda] \times [1/\tau, \tau]$. Then, we have

$$|\pi(\beta, \sigma \mid \mathbf{y}_n) - \pi(\beta, \sigma \mid \mathbf{y}_k)| = \pi(\beta, \sigma \mid \mathbf{y}_k) \left| \frac{\pi(\beta, \sigma \mid \mathbf{y}_n)}{\pi(\beta, \sigma \mid \mathbf{y}_k)} - 1 \right| \rightarrow 0 \text{ as } \omega \rightarrow \infty.$$

□

Proof of Results (c) and (d). Using Proposition 1, we know that $\pi(\beta, \sigma \mid \mathbf{y}_k)$ and $\pi(\beta, \sigma \mid \mathbf{y}_n)$ are proper. Moreover, using result (b), we have the pointwise convergence $\pi(\beta, \sigma \mid \mathbf{y}_n) \rightarrow \pi(\beta, \sigma \mid \mathbf{y}_k)$ as $\omega \rightarrow \infty$ for any $\beta \in \mathbb{R}$ and $\sigma > 0$, as a result of the uniform convergence. Then, the conditions of Scheffé's theorem (see Scheffé (1947)) are satisfied and we obtain the convergence in L_1 given by result (c) as well as the following result:

$$\lim_{\omega \rightarrow \infty} \int_E \pi(\beta, \sigma \mid \mathbf{y}_n) d\beta d\sigma = \int_E \pi(\beta, \sigma \mid \mathbf{y}_k) d\beta d\sigma,$$

uniformly for all rectangles E in $\mathbb{R} \times \mathbb{R}^+$. Result (d) follows directly.

□

Proof of Result (e). Using equation (1), result (e) follows directly from result (b).

□

References

References

- Andrade, J. A. A., O'Hagan, A., 2011. Bayesian robustness modelling of location and scale parameters. *Scand. J. Stat.* 38 (4), 691–711.
- Box, G. E. P., Tiao, G. C., 1968. A bayesian approach to some outlier problems. *Biometrika* 55 (1), 119–129.
- Chambers, R. L., 1986. Outlier robust finite population estimation. *J. Amer. Statist. Assoc.* 81 (396), 1063–1069.
- Desgagné, A., 2013. Full robustness in bayesian modelling of a scale parameter. *Bayesian Anal.* 8 (1), 187–220.
- Desgagné, A., 2015. Robustness to outliers in location–scale parameter model using log-regularly varying distributions. *Ann. Statist.* 43 (4), 1568–1595.
- Gwet, J.-P., Rivest, L.-P., 1992. Outlier resistant alternatives to the ratio estimator. *J. Amer. Statist. Assoc.* 87 (420), 1174–1182.
- Huber, P. J., 1973. Robust regression: Asymptotics, conjectures and monte carlo. *Ann. Statist.*, 799–821.
- Mallows, C. L., 1975. On some topics in robustness. Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.
- O'Hagan, A., Pericchi, L., 2012. Bayesian heavy-tailed models and conflict resolution: A review. *Braz. J. Probab. Stat.* 26 (4), 372–401.
- Peña, D., Zamar, R., Yan, G., 2009. Bayesian likelihood robustness in linear models. *J. Statist. Plann. Inference* 139 (7), 2196 – 2207.
- Scheffé, H., 1947. A useful convergence theorem for probability distributions. *Ann. Math. Statist.*, 434–438.
- West, M., 1984. Outlier models and prior distributions in bayesian linear regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 46 (3), 431–439.